

Relative Entropy Minimizing-Based Theory of Intelligent Systems¹

“ICAI: LATE BREAKING PAPER”

Guangcheng Xi

(State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, P.R.China)

Abstract: Based on the point of view of neuroethology and cognition-psychology, general frame of theory for intelligent systems is presented by means of principle of relative entropy minimizing in this paper. Cream of the general frame of theory is to present and to prove basic principle of intelligent systems: entropy increases or decreases together with intelligence in the intelligent systems. The basic principle is of momentous theoretical significance and practical significance. From the basic principle can not only derive two kind of learning algorithms (statistical simulating annealing algorithms and annealing algorithms of mean-field theory approximation) for training large kinds of stochastic neural networks, but also can thoroughly dispel misgivings created by second law of thermodynamics on *people's* psychology, hence make one be fully confident of facing life. Because of Human society, natural world, and even universe all are intelligent systems. In particular, Population systems are intelligent systems. According to the basic principles of intelligent systems, The intelligence of population system is proportional to the volume of population space, and decreases with the logarithm of population density..

Key words: Complex systems; Intelligent systems; Relative entropy minimizing; General frame of theory; Basic principle.

1 Introduction

Until now, human brain is the sophisticated creation of natural world and is only place where are produced cognition and intelligence, and hence spirit. Human being as a body carrying life is the most advanced and perfect intelligent system on the planet. It has the features common to general complex system. As we see it[1], a complex system is a functional system having any structure(including hierarchical and variably hierarchical structure) and consisting of any number of subsystems(≥ 1)capable of particular functions. These subsystems form closed loops of their own based on the feedback mechanism and are capable of self-adjustment according to different optimum criteria and interact with each other in various ways, and are composed of various dynamical, logical, nonlogical, and heuristic links. A complex system integrates at least the following six aspects:

- (1) Multi-dimensionality: magnitude of dimension, type of dimension, and transition each other of the types of dimension;
- (2) Multi-parameter;
- (3) Multi-relationship: relationships between(among)variables at same level(in particular,

¹ Having revisited content of the paper “Intelligent control with relative entropy minimizing (Control theory and applications, Vol.16, No.1, PP—27-31. In Chinese.)”, I have proposed and proved the basic principle of intelligent system. And then further having supplemented and perfected, this article is formed. It is submitted first time in English.

nonlinear relationships) and relationships between(among)variables at different level(in particular, nonlinear relationships)or the crossover of the two relationships;

(4) Multi-criterion: multi-component, multi-scales(levels);

(5) Multi-functionality: emergence, co-operation, competition, adaption, closed-openness, and so on;

(6) Multi-discipline of knowledge used.

Obviously, an intelligent system is complex system.

Intelligent control is the process of performing a task by an intelligent machine[2]. The theory of intelligent control is a composition of mathematic, lingual method and algorithm used in the system and the process and is based on cross disciplines of neuroethology, cognition-psychology, computer science, systematic science, artificial intelligence and information science. Only when we are accustomed both to introduction and effective use of all knowledge of modern science , in particular, of the systematic science in scientific fields related to “human being” and to introduction and effective use of all scientific knowledge related to “human bring” in other modern science, in particular ,in the fields of systematic science, in other words, only when human being has full knowledge of itself, can an intelligent control system in true sense be obtained and can an the theory of intelligent system in true sense be established.

Based on the point of view of neuroethology and cognition-psychology, general frame of theory for intelligent systems is presented by means of principle of relative entropy minimizing in this paper. Cream of the general frame of theory is to present and to prove basic principle of intelligent systems: entropy increases or decreases together with intelligence in the intelligent systems. The basic principle is of momentous theoretical significance and practical significance .From the basic principle can not only derive two kind of learning algorithms (statistical simulating annealing algorithms and annealing algorithms of mean-field theory approximation) for training large kinds of stochastic neural networks,but also can thoroughly

dispel misgivings created by second law of thermodynamics on *people's* psychology ,hence make one be fully confident of facing life.Because of Human society, natural world, and even universe all are intelligent systems.Human intelligence is in the Brain, intelligence of universe is in Black holes. In particular,highest intelligence of universe is in the black hole which possesses maximum volume in all black holes.

This paper is organized as follows:Section 2 presents the basic viewpoint and basic construction on intelligent systems,gives block diagram of intelligent control system.Section 3 presents and proves basic principle of intelligent system: entropy increases or decreases together with intelligence in the intelligent system.Section 4 demonstrates how derive two kind of learning algorithms (statistical simulating annealing algorithms and annealing algorithms of mean-field theory approximation) for training large kinds of stochastic neural networks by means of the basic principle of the intelligent system.

2 Basic viewpoint and basic construction on intelligent systems

The following is a block diagram of intelligent system from the perceptive of neuroethology and cognition-psychology.

It is a description of an abstracted framework of common intelligent control system, including the description of the hierarchical intelligent control system. We will give a brief description of the block diagram of intelligent control system shown in figure 1.

The perception system makes characteristic abstraction, transformation, integration and coding of environmental stimuli. The coded environmental stimuli are input to the memory system. The memory is categorized as short-term memory and long-term memory. Information processing by short-term memory comes from two sources. On one hand, it processes the coded information from environment. At this time, the short-term memory shows up great selectivity: elimination of the unwanted or minor information and preservation and further acknowledgment of the needed information. On the other hand, the short-term memory refines some portion of the information stored in the long-term memory. The long-term memory is a huge information storage-base storing various kinds of information, such as sport skills, grammar information, semantic information, pragmatic information, value, processing procedure, etc. Migrating with current and past input, some portion of the long-term memory is activated. The activated information is called active memory [3]. Some portion of the active memory undergoes a refining process in the short-memory which is the place where current cognitive activities take place. It is obvious that there must be a channel for information exchange between the short-memory and long-memory. The information exchange at the same time serves as a coordinator between the control system and reaction system. The control system itself has complete, multi-functional, high-level learning system. The control system determines how the entire intelligent control system works. By means of direct learning from the environment and making use of the information stored in the memory system, the control system implements its organization and management function, processes the objectives and provides policy to achieve these objectives, that is, determines a plan and make a decision. The coordination system is an intermediate structure between the control system and other subsystems and makes decision and coordination based on short-term memory, "to pass information from the higher level down to the lower level" and "to report information from the lower level to the higher level". The reaction system controls the output of the entire intelligent system ranging from various actions to language and expression.

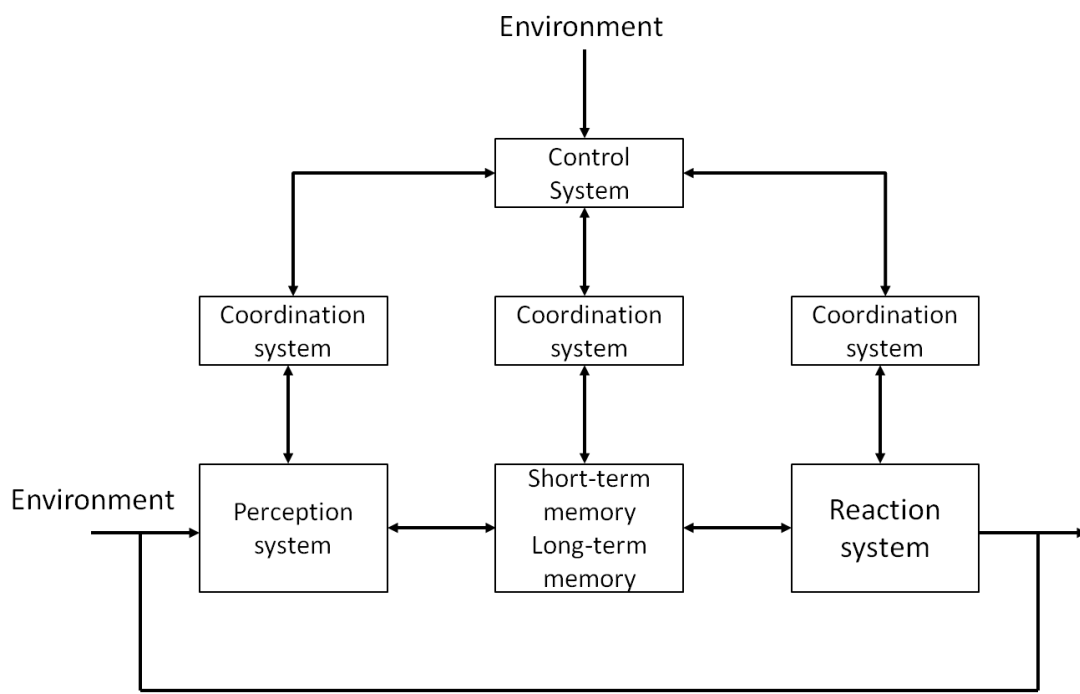


Fig.1 Block diagram of intelligent control system

3 Basic principle in intelligent system

Assume the intelligent system S shown in Figure 1 can be expressed as following triplet with a function

$$S = (X, F, P, H(p_0, p)) \quad (1)$$

Where X is the state space; F is a σ -algebra; P is a probability measure on (X, F) ; $H(p_0, p)$ is relative entropy defined as

$$H(p_0, p) = \int p_0(x, \theta) \ln \frac{p_0(x, \theta)}{p(x)} dx \quad (2)$$

Where $p_0(x, \theta)$ is determined by observation, $\theta = (\xi, \eta, t)$ is a parameter, where

ξ is the action rule of the intelligent system S , η is the event database, i.e. event set stored in the memory system, and t is time; $p(x)$ is a function of maximum entropy probability density representing a priori knowledge and is given by the following equation

$$p(x) = \frac{1}{Z} \exp(-mU(x)) \quad (3)$$

Where $U(x)$ is a vector potential, m is a vector factor and Z is partition function.

Whenever there is a new sufficiently large input to the system, the system executes a minimization process of formula (2) and when the process ends, the 0–1 symbol string supplied by the reaction system is the optimal task of the system.

The principle of minimum relative entropy is well known. And here, just to give

you a sense of what we're seeing.

Theorem 1. Express the relative entropy in equation (2) as $H(p_0, p)$, we have

$$\lim_{x \rightarrow \infty} H(p_o(x_i), p(x_i)) = 0. \quad (4)$$

Proof Because of $p_0(x_i)$ is estimates to $p(x_i)$ which are obtained by any practicable method to do random experiment (learning) about certain constraint of $p(x)$, for example, because of formula (3), there is a Markov chain of observable sample x_1, \dots, x_n whose limit distribution is $p(x)$ on state space X . By means of ergodic theorem, from orbits $x_i(t), i = 1, 2, \dots; t \geq 0$; of the Markov chain, estimate $p_0(x_i)$ of density function $p(x_i)$ can be obtained. The process is producing directive

sequence $p_0(x_i)$ ($i = 1, 2, \dots$) and its directive subsequence

$p_0(x_{L(i)}), L(i) > i, i = 1, 2, \dots$ on complete metric space (D, d) , D is set of probability density function on measurable space (X, F) . At this time,

$$\lim_{i \rightarrow \infty} p_0(x_i) = \lim_{i \rightarrow \infty} p_0(x_{L(i)}) = \lim_{i \rightarrow \infty} p(x_i).$$

In fact, suppose $\lim_{i \rightarrow \infty} p_0(i) = \bar{\eta}$, then for any small positive number ε , there is positive integer I , such that when $i > I$, $\|p_0(x_i) - \bar{\eta}\| < \varepsilon$. Because of $\lim_{i \rightarrow \infty} \{L(i)\} = \infty$, thus there is a positive integer i_0 , such that when $i > i_0, L(i) > I$, at this time $\|p_0(x_{L(i)}) - \bar{\eta}\| < \varepsilon$, hence subsequence $\{p_0(x_{L(i)})\}$ converges, and $\lim_{i \rightarrow \infty} p_0(x_i) = \lim_{i \rightarrow \infty} p_0(x_{L(i)}) = \lim_{i \rightarrow \infty} p(x_i) = \bar{\eta}$. The proof is completed.

The proving process of above theorem is the process in which $H(p_0, p)$ is minimized. The limit theory dictates that there must be a minimum value approaching to 0, and the minimum value is unique. The minimum value corresponds to the uniquely correct 0,1 character string for implementation of a given objective (environmental input). At control level, a 0,1 character string corresponding to the correct decision-making is given, and at the reaction system (level), a 0,1 character string corresponding to correct task is given.

From the above discussion, we find the following proposition

Proposition The sufficient and necessary condition for the intelligent system $S=(X,F,P,H(p_0,p))$ to exist is that it can be given by

$$S=(X,F,P,H(p_0,p))=\lambda \min_{p_0} \{ \int_X p_0(x,\theta) \ln \frac{p_0(x,\theta)}{p(x)} dx \}, \quad (5)$$

where $\lambda > 0$.

Definition 1. Knowledge on the intelligent system S is defined as the structural information on the intelligent system S , measured by equation (2).

Definition 2. The intelligence of an intelligent system is change ratio of the knowledge acquired by the intelligent system to the linkage coefficient between component units in the intelligent system.

Theorem 2 (Basic principle of intelligent system) . Entropy increases or decreases together with intelligence on the intelligent system S .

To make the discussion simple and sufficient, the concept of sufficiency of relative entropy is introduced [4].

Definition 3. For the intelligent system S , we assume \mathfrak{A} is σ -algebra on set \mathcal{X} of sub-vectors of $x \in X$ produced by its subsystem—memory system and control system, it is σ -subalgebra of F , that is, $\mathfrak{A} \subseteq F$. Let $P(X)$ is set of the all probability measures on (X,F) and $P' \subset P(X)$. By σ -subalgebra \mathfrak{A} is sufficient with respect to P' is meant there exists \mathfrak{A} -measurable function

$$h = E_\mu(1_A | \mathfrak{A}), \mu-a.e., \forall \mu \in P' \quad (6)$$

for any $A \in F$.

Assume P_0 and P are probability measures corresponding to probability density function $p_0(x)$, $p(x)$, respectively, $\{P_0, P\} \in P(X)$. From the definition of sufficiency we easily find that \mathfrak{A} is sufficient with respect to $\{P_0, P\}$. Therefore we have:

$$H_{\mathfrak{A}}(p_0, p) = H_F(p_0, p) = H(p_0, p).$$

Since the \mathfrak{A} is sufficient with respect to $\{P_0, P\}$, then the difference between P_0 and P can be determined all by \mathfrak{A} alone. Therefore the conclusion concerning

system $S = (X, F, P, H(p_0, p)) = \lambda \min_{p_0} \{ \int_X p_0(x, \theta) \ln \frac{p_0(x, \theta)}{p(x)} dx \}$ can be discussed

on its subsystems—memory system and control system.

Following this thinking in the discussion, we can not only prove the basic principle of the intelligent system S , but also produce a special algorithm—algorithm of mean-field theory approximation [5]. The proof of theorem 2 is given in following

Note that vector m in equation (3) denotes linkage coefficient between component units (nerve cells). For simplicity, we rewrite $p_0(x, \theta)$ as $p_0(x)$. We will find

$\partial H_{\mathfrak{A}}(p_0, p) / \partial m_i$ in the following.

$$\frac{\partial H_{\mathfrak{A}}(p_0, p)}{\partial m_i} = - \int \frac{p_0(x_T)}{p(x_T)} \frac{\partial p(x_T)}{\partial m_i} dx_T \quad (7)$$

Where $x_T \in \mathcal{X}$. We have:

$$\begin{aligned} \frac{\partial p(x_T)}{\partial m_i} &= [\exp(-\sum_i m_i U_i(x_T))] U_i(x_T) / Z_N - p(x_T) [\sum_{x_T} \exp(-\sum_i m_i U_i(x_T))] U_i(x_T) / Z_N \\ &= p(x_T) U_i(x_T) - p(x_T) \sum_{x_T} p(x_T) U_i(x_T) \end{aligned}$$

Where Z_N is restriction of Z on \mathcal{X} .

Substitute above equation into (7) we have

$$\begin{aligned} \frac{\partial H_{\mathfrak{A}}(p_0, p)}{\partial m_i} &= - \int \frac{p_0(x_T)}{p(x_T)} \frac{\partial p(x_T)}{\partial m_i} dx_T \\ &= - \left(- \int p_0(x_T) U_i(x_T) dx_T + \int p_0(x_T) \sum_{x_T} p(x_T) U_i(x_T) dx_T \right) \\ &= E_{p_0}(U_i(x_T)) - E_p(U_i(x_T)) \end{aligned} \quad (8)$$

Because of $E_{p_0}(U_i(x_T)) = H_{p_0} - \ln Z_N$, equation (8) reads

$$\frac{\partial H_{\mathfrak{A}}(p_0, p)}{\partial m_i} = H_{p_0}(x_T) - \ln Z_N - E_p(U_i(x_T)),$$

Consequently we have:

$$\ln Z_N + E_p(U_i(x_T)) = H_{p_0}(x_T) - \frac{\partial H_{\mathfrak{A}}(p_0, p)}{\partial m_i} \quad (9)$$

Where $H_{p_0}(x_T)$ is the maximum entropy determined by the observation sampling set or training sampling set, and is the measurement of uncertainty of task executed by the memory system and the control system. Notice $E_p(U_i(x_T))$ is determinate real number. After the number of component units (nerve cells) of memory system and control system is determined, Z_N (partition function) is determined as well. Therefore the left side of equation(9) can be regarded as a constant. If the first item at the right side increases (or decreases), the second item must increase (or decrease). Since \mathcal{A} is sufficient for $\{P_0, P\}$, the above conclusion is completely applicable to the entire system S . The proof of basic principle is completed.

Let's look at the population system. Population systems are intelligent systems. According to the basic principles of intelligent systems, **The intelligence of population system is proportional to the volume of population space, and decreases with the logarithm of population density..**

4 Intelligent behavior of intelligent system

Intelligent behavior of the intelligent system S is accomplished by the subsystems—control system and memory system. The control system and memory system are systems of neural networks. Various intelligent behaviors are shown biologically as the change of linkage weight between nerve cells.

Learning process is the process of acquiring knowledge. The learning process of system of neural network is divided into two phases. The first phase is a process in which the system of neural systems abstracts the environment and establishes the to be solved problem model—model establishment process. The second phase is a process in which the problem is solved—namely uncertainty caused by environmental regularity in the system is minimized. Corresponding to the two phases, there are the following results.

Theorem 3[6]. Assume the compounded system formed by system N of neural network and the environment E that it is in is isolated from the outside world, and the learning process of the system of neural network is a Markov process that has a given transition probability function. When input information of sufficient magnitude is obtained, the system of neural network starts its learning process. The process will finally arrive at its equilibrium from far from the equilibrium. The equilibrium is state when the system has maximum entropy. The sufficient and necessary condition of the maximum entropy is that the distribution density function of the system is given by

$$p(x_T) = \frac{1}{Z_N} \exp(-m_N U(x_T)) \quad (10)$$

Where x_T is state variable of the system of neural network, $U(x_T)$ is vector

potential function and Z_N is partition function.

Theorem 4[6]. Under the condition of the theorem 3, assume the linkage weight between the i th and j th nerve cells of system of neural network is w_{ij} and $w_{ij} = w_{ji}$, then based on basic principle of the intelligent system, we have

$$\frac{\partial H_{\mathfrak{A}}(p_0, p)}{\partial w_{ij}} = \gamma(P_{ij} - P'_{ij}) \text{ , where } \gamma = \frac{1}{kT} \text{ and } k \text{ is Boltzmann constant, } T$$

corresponds to the temperature in a physical system, but is a control parameter here; P_{ij} is the average probability when there is environmental input and the network arrives at equilibrium with both unit i and unit j conducting; P'_{ij} is the corresponding probability when there is not environmental input.

Theorem 5[5]. Under condition of theorem 3, assume that the linkage weight between the i th and j th nerve cells of system of neural network is w_{ij} and $w_{ij} = w_{ji}$, and that number of nerve cells of the system of neural network is sufficiently large, then based on the basic principle of the intelligent system and in the sense of mean square limit, we find $\frac{\partial H_{\mathfrak{A}}(p_0, p)}{\partial w_{ij}} = U_i(E_{p_0}(x_T)) - U_i(E_p(x_T))$, where E_{p_0} and E_p denote

expectation operators of $p_0(x)$ and $p(x)$, respectively.

From theorems 4 and 5 we find two kinds of learning algorithms for training stochastic neural networks:

$$\Delta w_{ij} = \beta(P_{ij} - P'_{ij}) \quad (11)$$

Where β is a constant less than but approaching 1.

$$\Delta w_{ij} = \alpha(E[\dot{x}_T(k)]E[\dot{x}_T(l)] - E[\bar{x}_T(k)]E[\bar{x}_T(l)]) \quad (12)$$

Where α is a constant less than but approaching 1, and $E[x_T(k)]$ is given by

$$E[x_T(k)] = \frac{1}{1 + \exp(\frac{-U_k}{T})}, \quad x_T(k) \in \{0, 1\} \quad (13)$$

Or

$$E[x_T(k)] = \tanh(\frac{-U_k}{T}), \quad x_T(k) \in \{1, -1\} \quad (14)$$

Where $U_k = \sum_l w_{kl} x_T(l)$; $E[x_T(k)]$ is mean field when unit k is excited, $E[\dot{x}_T(k)]$

and $E[x_T(k)]$ represent the excitement value of unit k when the system arrives at stable state in “-” phase(“minus” phase) and “+” phase(“plus” phase), respectively. “-” phase — with input unit to be clamping by environment(input mode), the excitement value of both hidden unit and output unit evolves to its stable value according to the rule expressed in equation(13) or (14); “+” phase — with the input unit yet to be clamping by environment(input mode) and the output unit to be clamping by the required output mode, only the hidden unit changes its excitement value till its stable value based on the same rule.

Directly applying learning algorithms (11) or (12) to the training of the system of neural network is exactly to use the fastest declining method to find the w_{ij}

when $H_{\mathfrak{A}}(p_0, p)$ approaching minimum, which will cause the system of neural system to be trapped into the local minimum, but in some cases, it is undesirable and sometime it is prohibited. Annealing algorithm is a method to avoid the local minimum and to obtain the overall minimum.

1)Statistic simulation annealing algorithm

Statistic simulation annealing algorithm[7] as a common method for approximately solving problem in large-scale combinatory optimization has seen great development in theory and practice. This algorithm can be widely used in the optimization field to provide solution to various optimization problems and the solution can arbitrarily approach to the overall optimization. This algorithm allow cost function related to state to climb the slope randomly simulating the metal annealing process. That is, with the function of the control parameter T , random noise is added to the conventional fastest decline process, thus if “unfortunately” the system of neural network runs into the local minimum trap, it will be able to get out of this local minimum trap until it maximally approaches the overall optimum, in other word, obtains the overall minimum of the cost function. In this article, the cost function is the relative entropy and the minimum process of relative entropy is synchronized with that of energy function—const function defined in other articles. The procedures for computer implementation of the statistic simulation annealing algorithm are:

- (1) Decide the sufficiently high initial T and randomly determine the linkage weight between nerve cells of the system.
- (2) Given a small perturbation to initial state x_T^i of the system, we find

x_T^i and the relative entropy increment $\Delta H_{\mathfrak{A}}$.

- (3) If $\Delta H_{\mathfrak{A}} \leq 0$, then accept this change, or otherwise if $\exp(-\Delta H_{\mathfrak{A}}) > \text{random number} \in [0, 1)$, then accept this change and $x_T^{i+1} = x_T^i$.
- (4) Compute new temperature: $T(i+1) = T(0) / \ln(i+1)$, $i \geq 1$, where $T(0)$ is the initial temperature.
- (5) Repeat steps (2) through to (4) until T approaches zero and the system no longer makes state transfer.

2) Annealing algorithm of mean-field theory approximation

The above statistic simulation annealing algorithm “inherits” the connatural essentiality of the Mote Carlo method —slow convergence rate and high computation complexity, which greatly restricts its application. For this reason, in recent years, specialists and scholars have put forward annealing algorithm of mean-field theory approximation, which has yielded good result. All methods for deducing annealing algorithm of mean-field theory approximation are based on the reduction of the system free energy in statistic physics. Our deduction[5] of this algorithm from the perceptive of relative entropy minimization is determined by the basic principle of the intelligent system and is a natural extension and result of the thinking—using relative entropy minimization to implement the intelligent control.

The annealing algorithm of mean-field theory approximation is as follows:

- (1) Randomly choose a high-value parameter T and randomly initialize the mean field $E[x_T(k)]$ and linkage weight w_{kl} of all free units.
- (2) Do following cycle :
 - a) Randomly choose a mean-field variable $E[x_T(l)]$, $l = 1, \dots, r$, and compute according to equation (13) or (14) till to arrive at stable state to find $E[x_T^1(1)]$.
 - b) Decrease T repeat step a) until a stable solution of equation (13) (or (14)) appears.
- (3) Implement the above procedures on “—” phase and “+” phase, respectively.
- (4) Modify weight according to equation (12).

The annealing algorithm of mean-field theory approximation has a convergence rate about 50~100 times higher than that of statistic simulation annealing algorithm and declines the time complexity from $O(2^n)$ to $O(n)$, where n is the number of nerve cells in the system of neural network.

The above two learning algorithm can be used to train large kinds of stochastic neural network, for example, Markov neural networks[6], neural network based on mean-field theory approximation[5], and so on. Our system of neural network in the intelligent system has learning function[6], memory function[8] and thought

function[9]. Having constructed the intelligent system ,having given mathematical method analyzing it and its algorithm, we have constructed theory of intelligent system, which is implemented through relative entropy minimization.

Reference

- [1]Guangcheng, X. (1987), “Entropy—method of partition of complex system” ,ACTA AUTOMATICA SINICA (In Chinese) ,Vol. 13, No. 3, pp. 216—220.
- [2]Mode M C and Saridis G N. (1990), “A Boltzmann machine for the organization of intelligent machines” , IEEE Trans. Syst., Man, and Cybern., Vol. 20, No. 5, pp. 1094—1102.
- [3]Lindsay PH and Norman DA. (1977), “HUMA INFORMATION PROCESSING An Introduction to Psychology” (Second Edition) ,Academic Press.
- [4] Halmos P R and Savage L T. (1949) “Application of the Radon—Nikodym theorem to the theory of Sufficient Statistics” , Ann. Math. Statistics ,Vol. 20, pp. 225—241.
- [5] Guangcheng, X. (1995), “Neural network Based on Mean-field theory approximation”, ACTA ELECTRONICA SINICA(In Chinese), Vol. 23, No. 8, pp. 62—64.
- [6]Guangcheng, X. (1991), “A tentative investigation of the learning process of neural network system” , ACTA AUTOMATICA SINICA(In Chinese), Vol. 13, No. 3, pp. 311—316.
- [7] Kirdpatrick S, Gellat C D and Vechi M P. (1983), “Optimization by Simulation Annealing” , Science , Vol. 220, No. 4598, pp. 671—680.
- [8] Guangcheng, X. (1998b), “A stochastic theory of associative memory” ,Control Theory and Application(In Chinese), Vol. 15, No. 5, pp. 688—694.
- [9]Guangcheng, X. (2003), “Variability of structure of Abstract Neural Automata and the ability of thought” , Kybernetes, Vol. 32, No. 9, pp. 1549—1554.

Acknowledgements

The research has been supported by National Basic Research Program of China (973 Program) under grant No. 2003CB517106.